Investigating Important Factors Influencing Iowa Alcohol Culture

Final Report for the Nonprofit Group: Drinking Excess Alcohol is Dangerous (DEAD)

Dylan Li, Alex Arrieta, Cameron Stivers, Ryuhei Shida, Nathan Hill

Introduction

Understanding the factors that drive alcohol purchases is vital to identifying and mitigating the risks inherently associated with alcohol culture. Recognizing the times and locations where increased risk occurs is essential to lowering risk. Within the state of Iowa, each of its 99 counties have differing characteristics. In order to identify the factors that pose the highest risk, the recent trends and patterns throughout the state must be exposed. Revealing factors which are correlated with changes in alcohol sales will be vital in future attempts to promote safer alcohol culture.

Data Preparation

In order to begin determining factors associated with increased alcohol consumption in Iowa, we utilized two different data sources. Our main source of data came from the Iowa Department of Revenue, Alcoholic Beverages[1]. This dataset contained all class "E" liquor sales in the state from the beginning of 2012 through the first nine months of 2023. The information we were primarily interested in from this source included the volume and type of alcohol, date, location, type, and vendor for each purchase. One of the challenges involved with this data is the fact that it records the store purchases of alcohol exclusively. Thus, we must assume that alcohol purchased by the store generally reflects the patterns of sale and consumption around the time of purchase. In order to account for this vague time period, we became interested in grouping the alcohol sales by month rather than by day or week.

The second source of data that we hoped to investigate for analyzing trends in Iowa liquor sales is population data from the United States Census Bureau[2]. The populations are broken down by individual counties so that we could analyze the average amount of alcohol purchased by an individual person in a given location. This approach makes interpretations and results much more tangible than dealing with total amounts of alcohol sold. We also divided the counties into five groupings based

1 https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy

2 https://www.iowa-demographics.com/counties_by_population

on population size, putting an equal number of counties in each group. This way, we were also interested in whether the population size of a county has any effect on the amount of alcohol purchased there.

We chose to use information only from 2020 through 2022 based on summary information that indicated there is very little variation in the volume of alcohol purchased from year to year. In comparison, the seasonal variation for alcohol purchases within a year appears relatively high. Thus, looking at the three most recent full years of alcohol sales is adequate for revealing trends. All in all, we utilized 7,733,863 individual sales to investigate this issue, grouping by month and county.

Model Selection

The main factors that we were interested in investigating were month, year, county population, category of alcohol, and vendor. We quickly realized that the year of purchase is much less significant in predicting volume of alcohol than it is in predicting price. We primarily focused on maximizing the R-Squared metric while manufacturing a response that would be meaningful to interpret. While “category” and “vendor” boosted our metrics, these categories did not necessarily help us in understanding the factors that drive alcohol culture. The major insight gained from these factors is that different types of alcohol are sold in vastly different volumes. Thus, the month of the alcohol purchase and the population of the county where it was purchased became our two primary factors for uncovering patterns in alcohol sales. We fit our data into a five-fold cross validation model with ridge regression. We tested a multitude of ridge penalties, including an ordinary least squares model with no penalty, using R-Squared as our primary metric. After this testing, the final ridge penalty that maximized the R-Squared value of the model was 0.5.

Final Model

Our final model uses the predictors of month and size of county by population to predict the amount of alcohol purchased by an average resident in a given county in a certain month. Using a ridge regression with a penalty of 0.5, the resulting R Squared value of 0.237 indicates that 23.7% of the variation in the average amount of alcohol purchased in a county per month can be explained by the population size of the county and the month of the year. The intercept value for this model is 0.136. This value can be interpreted as the number of gallons of alcohol an individual resident of a very large county in Iowa would

be expected to purchase in the month of January.

Factor	Value
Intercept	0.13614035
February	0.01196030
March	0.02354895
April	0.03039265
May	0.04142543
June	0.04697162
July	0.03838492
August	0.03817008
September	0.03445687
October	0.03140704
November	0.03298418
December	0.05237140
Large Counties	-0.02571957
Medium Counties	-0.04230852
Small Counties	-0.05922209
Very Small Counties	-0.08441690

Results

Our model illustrates the patterns both in month and in county size that impact Iowa alcohol consumption. Firstly, it is clear that all of the month estimators in the model are positive, thus January has the lowest alcohol sales statewide. Interestingly, the largest month for alcohol sales by volume tends to be December. Based on our model, we estimate that 0.0524 more gallons of alcohol are purchased per resident in December than in January after adjusting for the size of the county. This amounts to nearly a cup per person, making it quite significant in the overall image of alcohol culture throughout the state.

Furthermore, the largest 20% of counties in Iowa tend to sell larger quantities of alcohol per person relative to ones with smaller populations. In fact, this trend continues, as large counties, medium counties, and small counties sell less and less alcohol per resident until the smallest 20% of counties sell the least per person. Relative to the very large counties, the very small counties expect to sell 0.0844 gallons less per resident every month, after adjusting for the month. The counties in the top 20% by population in Iowa are Polk, Linn, Scott, Johnson, Black Hawk, Dallas, Woodbury, Story, Dubuque, Pottawattamie,

Warren, Clinton, Cerro Gordo, Muscatine, Marshall, Des Moines, Jasper, Webster, and Sioux Counties.

Recommendations

Based on the results provided by our model, we recommend concentrating efforts to make alcohol culture safer in Iowa in the months of December and June specifically in larger counties. These efforts will be able to impact the largest number of people, in the areas where alcohol purchases per person are the highest, at the time when alcohol purchases per person are highest. Some possible explanations for the emergence of these patterns include holiday seasons, such as Christmas and New Year's in December and Fourth of July immediately following June. It also makes sense that larger counties have higher alcohol purchases per person, since they are the counties that are more likely to attract visitors and there are more opportunities to purchase alcohol. Overall, focusing specifically on these areas during the most active times of year will allow an effective campaign to make a positive difference in Iowa's alcohol culture.

Ethical Considerations

The data used in this process was publicly available data published by the State of Iowa. All usage of this data and model should follow Iowa State Government guidelines found here.[3] Additionally, despite this information being public, it should not be used against any of the companies specifically listed in the data directly or indirectly in manners such as attempting to slander these companies. Furthermore, the exact patterns provided in this report are not guaranteed to hold true in the future, they are simply an explanation of an existing pattern that could change due to new circumstances. Finally, although we often reference the "average resident" in a certain county, we can not assume anything about individual residents from the model and summaries. The conclusions we have come to reflect the county as a whole, and judgment should not be passed to individuals during efforts to make alcohol culture safer.

[3] https://www.iowa.gov/policies

Tables and Visuals

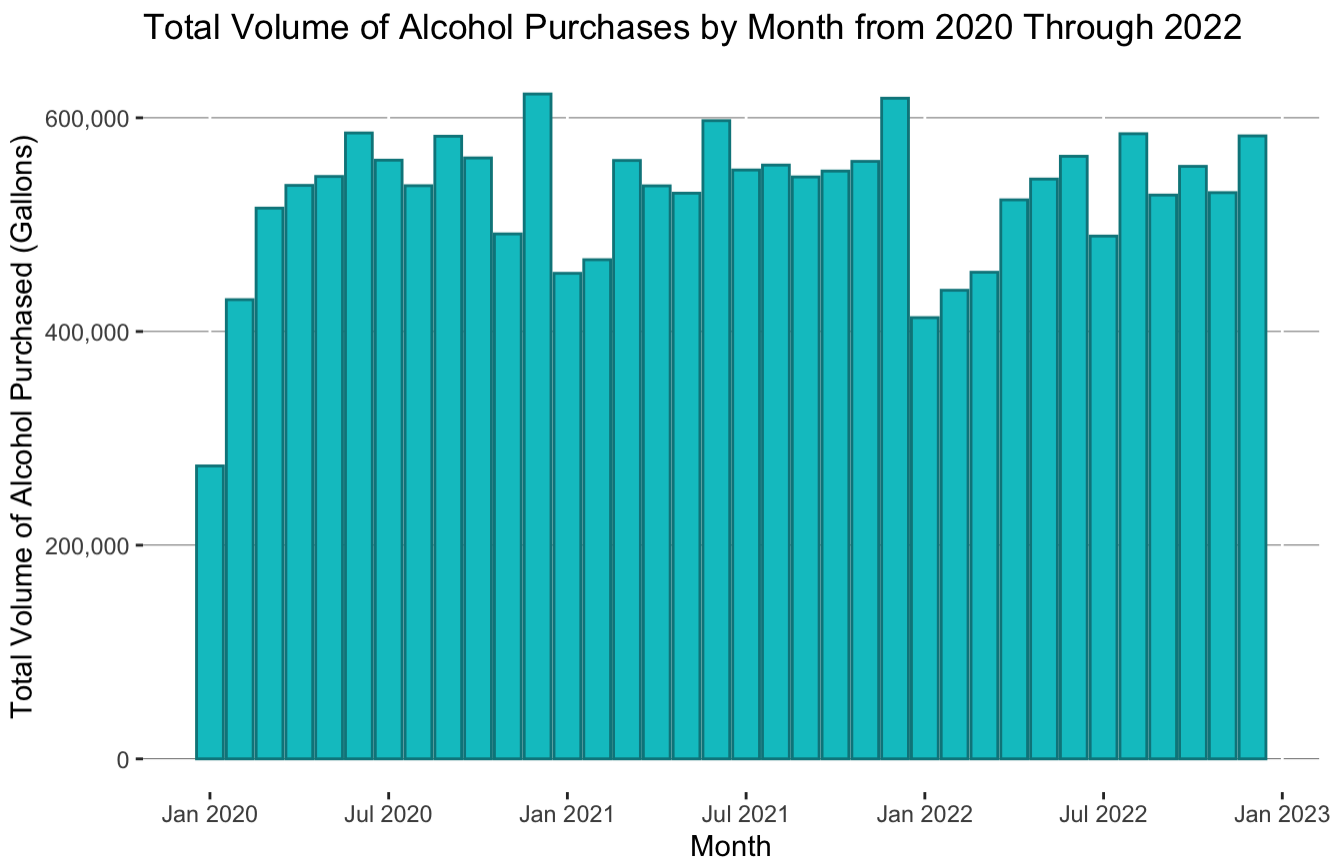

Figure 1: Seasonal trend of alcohol purchases by volume throughout Iowa. The early winter months of the year yield less alcohol sales by volume compared to the other months in 2020, 2021, and 2022.

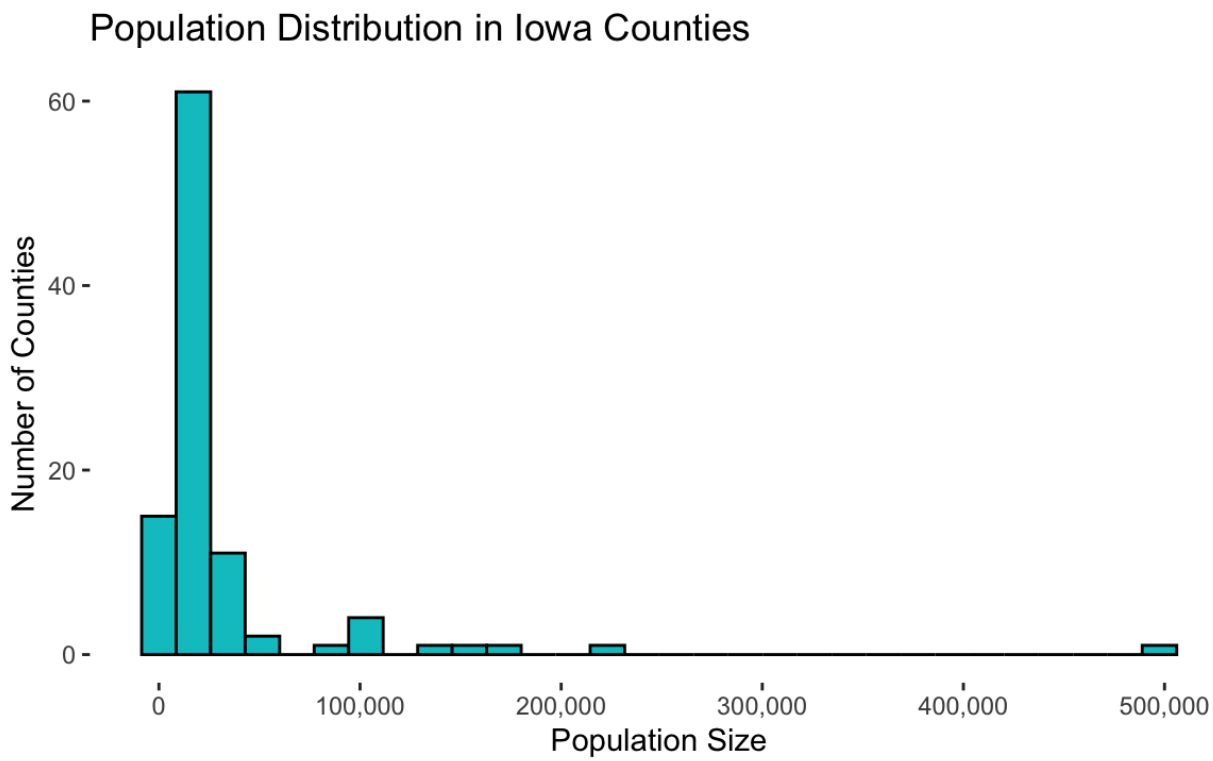

Figure 2: Population Distribution of Iowa Counties. The largest county in Iowa by population is Polk.

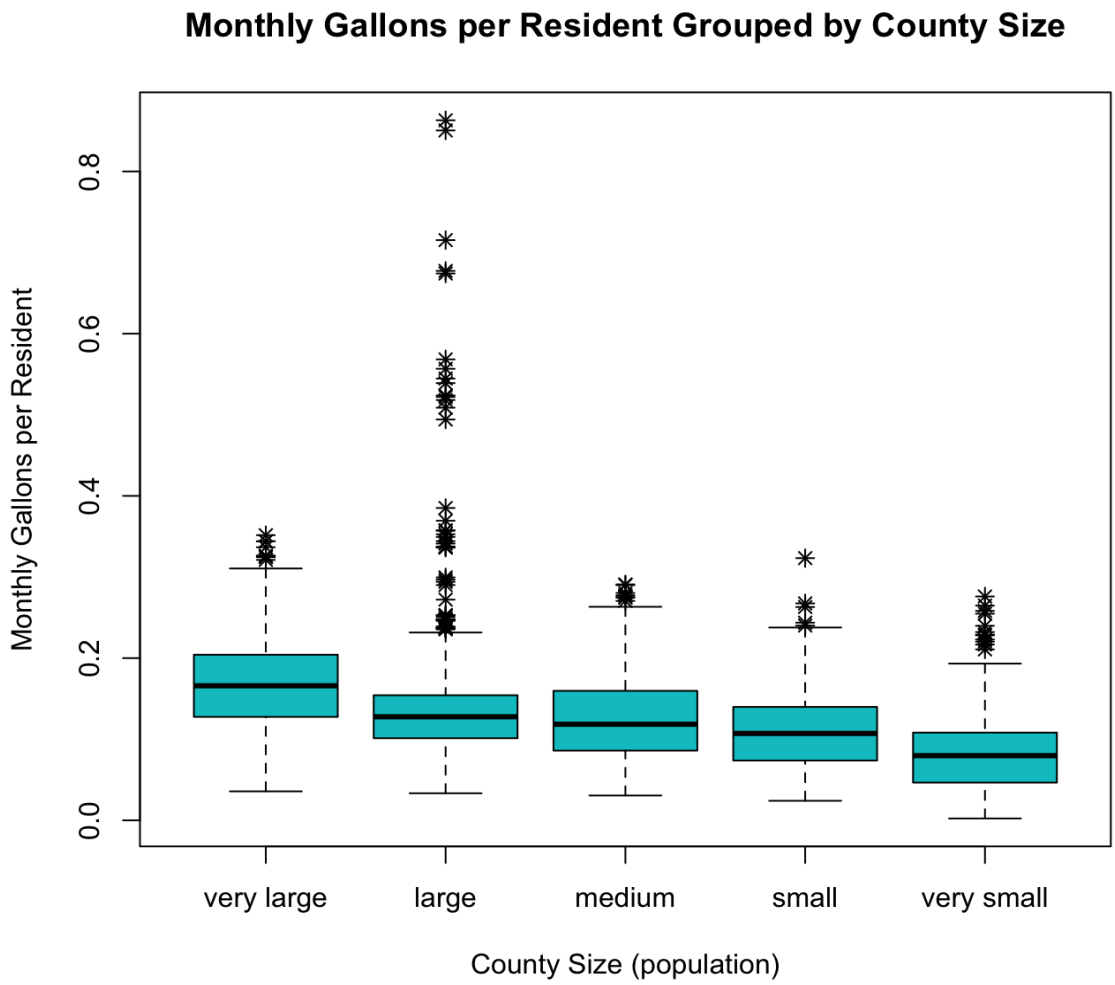

Figure 3: This boxplot portrays the trend associated with alcohol purchases in counties organized by population size. Clearly, the average large counties generally involve higher volumes of alcohol purchased per person relative to smaller counties.